VALUE IN HEALTH 16 (2013) 434–437

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

METHODOLOGICAL ARTICLES

An Expert on Every Street Corner? Methods for Eliciting Distributions in Geographically Dispersed Opinion Pools

Daniel Sperber, MSc[1,*], *Duncan Mortimer, PhD*[1], *Paula Lorgelly, PhD*[1], *David Berlowitz, PhD*[2,3,4]

[1]*Centre for Health Economics, Monash University, Melbourne, Australia;* [2]*Institute for Breathing & Sleep, Melbourne, Australia;* [3]*Austin Health, Melbourne, Australia;* [4]*University of Melbourne, Melbourne, Australia*

A B S T R A C T

Recent publications outline developments in eliciting probabilistic opinions from clinical experts with which to inform structural assumptions and parameter estimates in health economic models. We outline approaches taken to date to elicit probabilistic distributions from experts within the health economic literature and outline the appropriate considerations and the resulting process in developing a new elicitation program with the aim of allowing low-cost elicitation of expert opinion from a heterogeneous and geographically dispersed opinion pool while preserving the essential features of good practice elicitation methods.

Keywords: clinical opinion, elicitation, expert opinion.

Introduction

Recent "best-practice" guidelines in decision analytic modeling [1] suggest that clinical opinion may be included in economic models where necessary and well documented, despite broad concerns over its nonexperimental nature, potential overconfidence, and general susceptibility to biases and heuristics in memory and decision making. If clinical opinion is to be used influentially in decision models, further examination of elicitation methods may be appropriate, especially around the elicitation of uncertain probabilistic parameters. A small literature has developed within the field regarding the direct elicitation of probability distributions from clinicians and clinical researchers [2–5]). Existing examples of good practice such as the Sheffield Elicitation Framework (SHELF) [6] rely on personal interview and are therefore of limited assistance in eliciting opinion from geographically dispersed opinion pools. Our elicitation exercise seeks to build upon existing applied examples by attempting to emulate aspects of SHELF in a spreadsheet program, thereby allowing for wider dissemination and completion, to inform the economic evaluation of a multicenter clinical trial in a rare treatment area: sleep apnea related to acute quadriplegia.

Methods

A targeted review of elicitation methods in health economic applications was undertaken. We also reviewed guidelines for elicitation methodology including the psychology of elicitation, visual aids, and statistical methods for probabilistic elicitation and response aggregation. SHELF [6] was identified as an example of good practice, having been recently developed by an experienced group of elicitation methodologists in concert with an elicitation textbook [7] and a series of freely available research tools and programs.

It was not clear, however, that the personal interview program embodied in the SHELF recommendations was appropriate for our potential respondent pool. In an effort to design an elicitation framework suitable for the collection of opinions from a geographically dispersed, heterogeneous, and very busy group of respondents, we emulated aspects of the SHELF framework where possible and incorporated necessary departures in the SHELF framework with respect to elicitation survey design, expert selection, calibration, and weighting. An Excel spreadsheet was developed, pilot tested, and is currently being disseminated to a population of researchers and clinicians.

Results

Literature Review

The literature review identified a number of existing protocols for expert elicitation [2–6]. The method adopted by Stevenson et al. [2] conforms closely to good-practice guidelines arising from the general probability elicitation literature [7–10] in which in-person

* *Address correspondence to:* Daniel Sperber, Room 278, Centre for Health Economics, Building 75, Monash University, Clayton, Victoria, Australia, 3800.

E-mail: Daniel.sperber@monash.edu.

1098-3015/$36.00 – see front matter Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2012.10.011

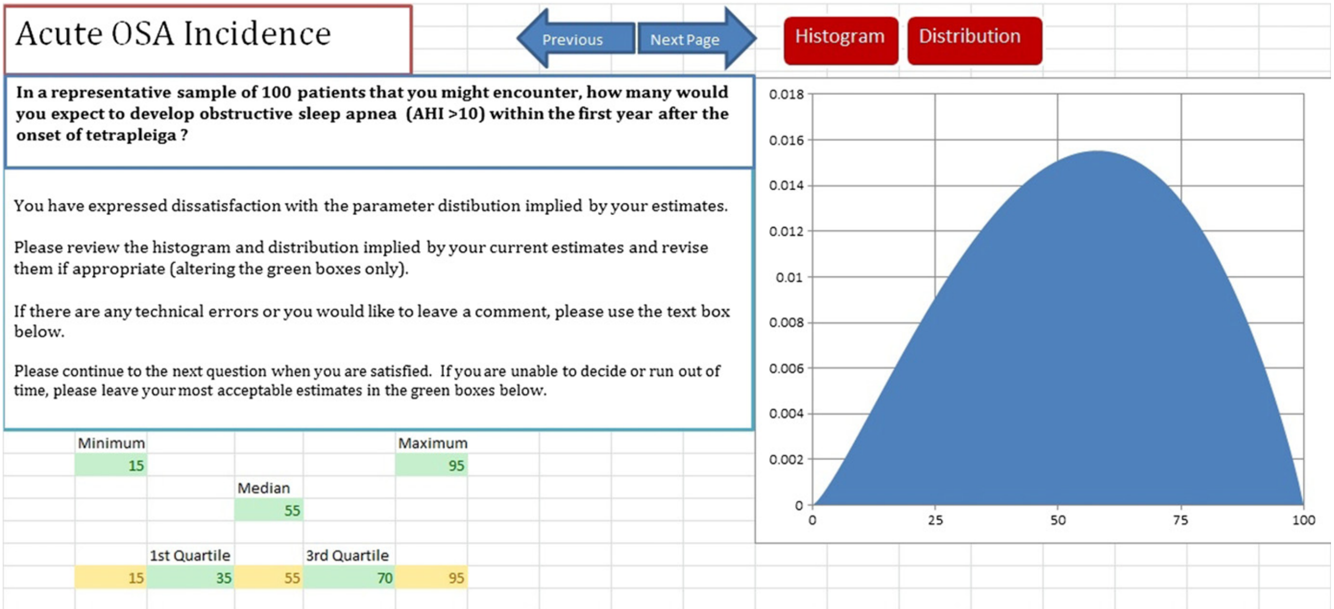

Fig. 1 – Example elicitation page. OSA, obstructive sleep apnea; AHI, Apnea–Hypopnea Index.

elicitation by an experienced facilitator is considered to be a first-best approach in eliciting an expert's true parameter estimate.

The major applied challenges with intensive, in-person elicitation programs relate to expert recruitment, which may be constrained by geography, accessibility, and resources. Restricting the opinion pool to a small sample of geographically proximate and highly committed collaborating experts—as is typically necessary for in-person elicitation—may amplify biases arising from local paradigms, self-interest, or shared exposure to unrepresentative clinical experience or a selected body of research among a group of peers. An online tool for the SHELF program has recently become available, which may provide another means of elicitation at a distance [11].

The "Excel spreadsheet" approach of Leal et al. [3], Bojke et al. [4], and Soares et al. [5] is engineered to encourage ease of use and would allow dissemination to a wider opinion pool. For example, Leal et al. and Bojke et al. e-mailed their survey to a relatively wide opinion pool of approximately 15 identified experts, resulting in a response rate of approximately 30% to 50% of the sample frame. Soares et al. used the spreadsheet format as an aid to a large group elicitation of local nurses. We considered that each of these existing Excel-based protocols lacked one or more elements of best practice (e.g., provision of visual and quantitative feedback, anchoring and adjustment errors, and appropriate aggregation methods) as described by O'Hagan et al. [7] and reflected in the SHELF framework [6] and Stevenson et al. [2].

In practice, it may be the case that there exists a trade-off between the ability of an elicitation exercise to survey the "true" personal belief of an expert and to be a palatable and feasible undertaking for a wide cross section of respondents.

Program Development

An Excel workbook was programmed to create an elicitation experience evoking that of a structured interview. Respondents begin with reading approximately 10 minutes of context: the rationale for the elicitation, an introduction to the elicitation process, and a summary of common biases and heuristics in decision making. Following a stylized trial elicitation and a calibration question, they are asked to estimate the probability distribution describing their uncertainty regarding two structural assumptions in our decision model.

A quartile-bisection approach was adopted to elicit respondent distributions and attempt to mitigate the impact of anchoring effects. Instantaneous visual feedback with regard to the impact of quartile estimates was provided by a histogram and beta distribution fitted by least squares using Easyfit software (Figure 1) [12].

The prepared elicitation program departed markedly from the SHELF process in conducting elicitation at a distance, by foregoing a group dynamic, and by diluting the implicit "expert" standard. As such, the program is not a SHELF implementation but an effort to emulate selected points of emphasis in the elicitation recommendations within a spreadsheet program. Direct adaptations include the quartile elicitation procedure, the emphasis on visual feedback to the respondent, and the introductory expert training material [13] (see Appendix 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2012.10.011).

Our elicitation plan sought to avoid the various problems associated with restricting the opinion pool to local collaborating experts by distributing a spreadsheet program widely: to published authors in our research area, international trial collaborators, and unaffiliated physicians identified via the Australian Medical Register. Given the lack of an in-person interviewer to dissuade irrational responses, the handling of potential irrational or faulty elicitation responses was of particular concern. In light of this concern, we tested a variety of response aggregation methods, the most established of which were an unweighted linear opinion pool and a weighted linear opinion pool informed by "best estimate" fractions [4] arising from Monte Carlo sampling of the calibration distribution against a relatively known parameter distribution (in our case, informed by the sole relevant clinical publication in the area) [14].

We also tested the behavior of a couple of arbitrary weighting methods: one including only our clinical collaborator, to reflect the impact of not performing a broader elicitation exercise on our model, and another using only the median estimates provided by respondents (which we assume to be reflective of the deterministic estimates they might have provided to another research design, and to which we fitted a monomodal beta distribution with the specified mean, and a maximized SD).

Finally, for the purposes of sensitivity analysis, we populated a weighted linear opinion pool on the basis of the qualitative judgments of our analyst, upweighting or downweighting responses to each question on the basis of whether the respondent had direct content area experience (double weight) and the positive relationship between quantitative and qualitative feedback (half weight for minor discrepancies, exclusion from the analysis for major plausibility issues). An example relating to overconfidence is described below.

Pilot Testing and Revision

Pilot testing was performed with local clinical collaborators. Respondent feedback was positive, but there was evidence of limited respondent interaction with the feedback and revision protocols. There were a small number of responses in which the quantitative elicitation methods did not appear to be effectively communicating the respondent's beliefs including one set in which qualitative statements of uncertainty (e.g., "I don't know, this isn't my area of research") were accompanied by extremely dense probability estimates (a 100% weighting between the values of 0 and 0.1), when one would hope that a very uncertain respondent would provide a very diffuse estimate.

A response that is representative of a second set of suspected errors is illustrated in Figure 2, in which the respondent has biased the first and third quartile values toward the outlier values, resulting in probability estimates for tail values that are greater than the median and mean. While distributions of this shape are certainly possible, we have no reason to believe that this was the intention of the respondent, or a plausible estimate of the underlying parameter.

To minimize these occurrences and provide evidence of their development through the exercise, the implicit choice of respondents of whether or not to review their responses graphically was removed. A recorded iterative progression from initial estimation to histogram feedback to the fitted distribution algorithm allows the examination of the impact of each marginal step on responses.

It remains to be seen whether the use of fitted distributions will offer sufficient value to justify the need for respondents to install new software, with a potential negative impact on the response rate. Likewise, we found little evidence of the superiority of one weighting method in our pilot data and would likely be satisfied with an unweighted opinion pool.

Discussion

There is good reason to proceed with caution in the adoption of elicited information in health technology assessment processes and in research prioritization decisions. In the presence of various known biases, it is unlikely that an accurately elicited estimate of clinicians' opinions will mirror the true underlying uncertainty of a treatment effect. Moreover, elicitation processes that are performed alongside health technology assessment programs may be expected to fall short of accurately eliciting the population estimate because of small, unrepresentative samples, or larger, more representative samples that have relatively poor preparation for the difficult task at hand.

The program described here was developed with the aim of allowing low-cost elicitation of expert opinion from a geographically dispersed opinion pool while preserving the essential features of recommended elicitation methods. The intent behind making such tools available is to facilitate the widespread collection of higher quality data, which may stand to provide potential advances in two major fields: the displacement of informal analyst assumptions or informal elicitation evidence and estimation of the value of future research relating to research questions that are not currently addressed.

Anyone who would like to discuss further applications of these methods is welcome to contact the corresponding author at daniel.sperber@monash.edu. A copy of the elicitation program is available as Appendix 2 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2012.10.011).

Source of financial support: This research was funded in part by the Victorian Transportation Accident Commission. The authors have no direct or indirect financial relationship with the sponsor.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2012.10.011 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

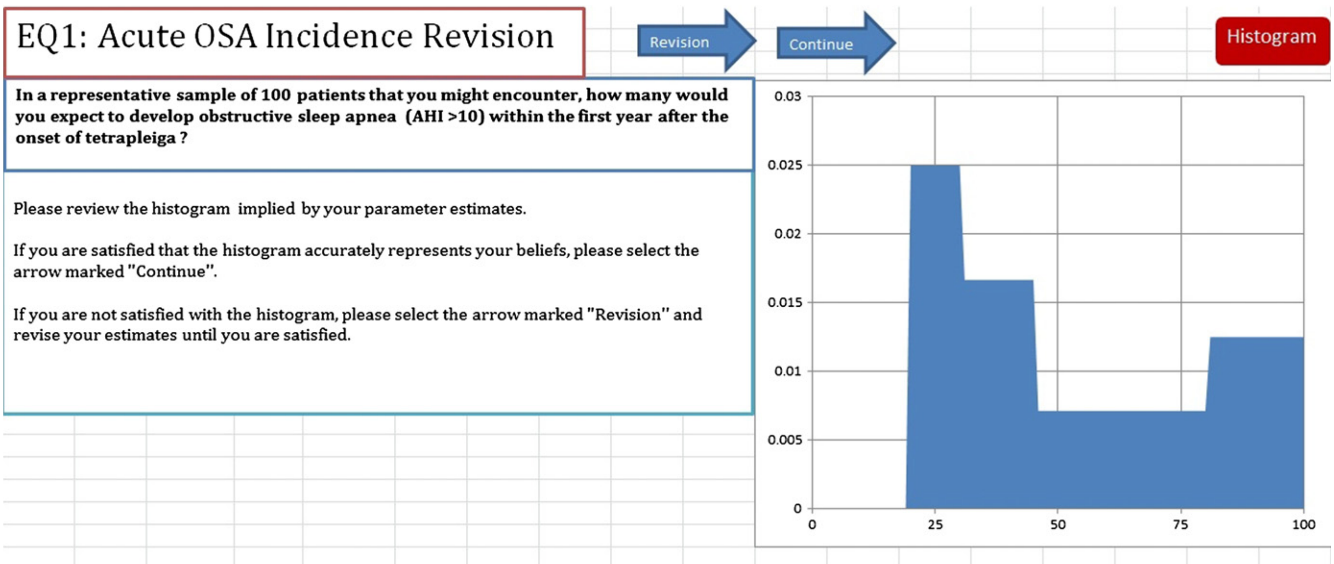

Fig. 2 – Sequential revision worksheets. OSA, obstructive sleep apnea; AHI, Apnea–Hypopnea Index.

REFERENCES

[1] Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices—modeling studies. Value Health 2003;6:9–17.
[2] Stevenson MD, Oakley JE, Lloyd Jones M, et al. The cost-effectiveness of an RCT to establish whether 5 or 10 years of bisphosphonate treatment is the better duration for women with a prior fracture. Med Decis Making 2009;29:678–89.
[3] Leal J, Wordsworth S, Legood R, Blair E. Eliciting expert opinion for economic models: an applied example. Value Health 2007;10:195–203.
[4] Bojke L, Claxton K, Bravo-Vergel Y, et al. Eliciting distributions to populate decision analytic models. Value Health 2010;13:557–64.
[5] Soares MO, Bojke L, Dumville J, et al. Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. Stat Med 2011;30:2363–80.
[6] Oakley J., O'Hagan A. SHELF: The Sheffield Elicitation Framework (version 2.0). Sheffield, UK: School of Mathematics and Statistics, University of Sheffield, UK, 2010 [updated 2010; cited]. Available from: http://tonyohagan.co.uk/shelf. [Accessed October 1, 2012].
[7] O'Hagan A, Buck CE, Daneshkhah A, et al. Guidance on Best Practice. Uncertain Judgements: Eliciting Experts' Probabilities. London: John Wiley & Sons, Ltd., 2006.
[8] Cooke RM. Experts in Uncertainty: Opinion and Subjective Probability in Science. Oxford, UK: Oxford University Press, 1991.
[9] Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. J Am Stat Assoc 2005;100: 680–701.
[10] Kynn M. The 'heuristics and biases' bias in expert elicitation. J R Stat Soc Ser A Stat Soc 2008;171:239–64.
[11] Morris E, Oakley J. MATCH Uncertainty Elicitation Tool. 2012 [updated 2012; cited]. Available from: http://optics.eee.nottingham.ac.uk/match/uncertainty.php#. [Accessed September 15, 2012].
[12] Mathwave. EasyFit. 5.5 ed, 2011. Available from: http://www.mathwave.com/ [Accessed May 10, 2011].
[13] Oakley J, ed. Eliciting Univariate Probability Distributions. London: Risk Books; 2010.
[14] Berlowitz D, Brown D, Campbell D, Pierce R. A longitudinal evaluation of sleep and breathing in the first year after cervical spinal cord injury. Arch Phys Med Rehabil 2005;86:1193–9.